

# Brain Stroke Prediction Using Machine Learning

<sup>1</sup>Lokesh Rathod, <sup>2</sup>Shailesh Bhole, <sup>3</sup>Nihar Bhakre

<sup>1,2,3</sup>Student, Artificial Intelligence and Machine Learning Engineering, Ajeenkya DY Patil School of Engineering, Pune, Maharashtra, India

**Abstract** - Stroke is a major global health concern and remains one of the leading causes of death and long term disability. Early prediction of stroke risk plays an important role in preventive healthcare and can significantly reduce mortality rates. With the increasing availability of healthcare data, machine learning techniques provide powerful tools for analyzing patient information and identifying risk patterns. This research presents a machine learning based approach for predicting the probability of brain stroke using patient health parameters such as age, hypertension, heart disease, body mass index, glucose level, and lifestyle factors. Various machine learning algorithms are explored and evaluated to determine their effectiveness in predicting stroke risk. The proposed system demonstrates how data driven models can support healthcare professionals in early diagnosis and preventive care.

**Keywords:** Stroke Prediction, Machine Learning, Healthcare Analytics, Predictive Modeling, Artificial Intelligence.

## I. INTRODUCTION

Stroke occurs when the blood supply to part of the brain is interrupted, leading to brain cell damage due to oxygen deprivation. According to global health reports, millions of people experience stroke every year and many of them suffer permanent disabilities. Traditional diagnosis methods rely on clinical observations and medical tests. However, predicting stroke risk before symptoms appear is difficult using conventional approaches. Machine learning provides the ability to analyze large datasets and identify complex relationships among medical variables. By training predictive models on historical patient data, machine learning algorithms can detect hidden patterns and estimate the likelihood of diseases. The aim of this project is to develop a machine learning based stroke prediction system that analyzes patient health attributes and predicts stroke risk.

## II. OBJECTIVES OF THE STUDY

- To analyze healthcare data related to stroke risk factors.
- To implement machine learning algorithms for stroke prediction.
- To evaluate model performance using classification metrics.

- To develop a predictive system that can assist healthcare professionals.

## III. LITERATURE REVIEW

Several researchers have explored machine learning applications in healthcare prediction. Logistic Regression, Decision Tree, Support Vector Machine, and Random Forest are commonly used algorithms for classification tasks. Previous studies show that parameters such as age, hypertension, diabetes, and lifestyle habits significantly influence stroke risk. Ensemble learning models often achieve better performance due to their ability to combine multiple decision trees. This research builds upon previous work by implementing different machine learning models and evaluating their predictive performance on stroke related datasets.

## IV. DATASET DESCRIPTION

The dataset used in this study contains patient health information and demographic attributes. Each record represents a patient with several medical parameters that may influence stroke risk.

Feature	Description
Age	Age of the patient
Gender	Male or Female
Hypertension	High blood pressure condition
Heart Disease	Presence of heart disease
BMI	Body Mass Index
Average Glucose Level	Blood glucose measurement
Smoking Status	Smoking habits
Stroke	Target variable indicating stroke occur

## V. METHODOLOGY

The proposed stroke prediction system follows a structured machine learning pipeline.

### Data Preprocessing

Missing values are handled and categorical variables are converted into numerical form.

Normalization techniques are applied to improve model performance.

### Feature Selection

Important medical parameters are selected to reduce noise and improve prediction accuracy.

### Model Training

Multiple machine learning algorithms are implemented including Logistic Regression, Decision Tree, and Random Forest.

### Model Evaluation

The models are evaluated using standard classification metrics such as accuracy, precision, recall, and F1 score.

## VI. MACHINE LEARNING ALGORITHMS USED

**Logistic Regression:** A statistical model used for binary classification problems.

**Decision Tree:** A tree based model that splits the dataset based on decision rules.

**Random Forest:** An ensemble method that combines multiple decision trees to improve accuracy.

## VII. RESULTS AND DISCUSSION

Experimental results indicate that machine learning models can effectively predict stroke risk based on patient health attributes. Among the tested models, Random Forest

achieved the highest prediction accuracy due to its ensemble learning capability. The analysis also revealed that age, hypertension, and glucose level have a strong influence on stroke probability. These findings highlight the importance of predictive analytics in healthcare decision making.

## VIII. ADVANTAGES OF THE PROPOSED SYSTEM

- Early identification of high risk patients.
- Data driven healthcare decision support.
- Reduced risk through preventive medical actions.
- Efficient analysis of large medical datasets.

## IX. CONCLUSION

This study demonstrates the effectiveness of machine learning techniques for predicting brain stroke risk using healthcare data. The developed system highlights how predictive models can support early diagnosis and preventive healthcare strategies. Future work may involve integrating deep learning models, larger medical datasets, and real time monitoring systems to improve prediction accuracy and clinical usefulness.

## REFERENCES

- [1] World Health Organization. Global Stroke Statistics.
- [2] Han, J., Kamber, M., & Pei, J. Data Mining: Concepts and Techniques.
- [3] Bishop, C. Pattern Recognition and Machine Learning.
- [4] Goodfellow, I., Bengio, Y., & Courville, A. Deep Learning.
- [5] Healthcare Analytics Research on Stroke Prediction using Machine Learning.

### Citation of this Article:

Lokesh Rathod, Shailesh Bhole, & Nihar Bhakre. (2026). Brain Stroke Prediction Using Machine Learning. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(3), 209-210. Article DOI <https://doi.org/10.47001/IRJIET/2026.103029>

\*\*\*\*\*